BRAIN-AGE PREDICTION

By
Yue Xu

Senior Thesis in Computer Engineering
University of Illinois at Urbana-Champaign

Advisors: Yogatheesan Varatharajah (Ph.D Student)
Professor Zbigniew T. Kalbarczyk

May 2019

Abstract

Brain age is a popular measure used in the study of brain aging that estimates the biological age of a brain based on the extent of cerebral atrophy. Accurate brain-age prediction models have wide applicability in the clinical domain, e.g., predicting age-related neurodegenerative diseases. Brain-age models are usually developed by learning the relationship between chronological age and brain structure in samples of healthy individuals with the reasoning that a healthy individual's brain age will be close to his or her chronological age. Furthermore, brain-structure-related features are derived based on structural magnetic resonance imaging (SMRI) data registered to standard atlases and extracting volumetric measures of standard brain regions. The primary goal of this project was to evaluate the efficacy of a brain-age-prediction model that used raw SMRI images instead of preprocessed features. A model that uses raw images to make brain-age predictions can eliminate the need for domain expertise in the development of predictive models, and also identify new biomarkers related to certain neurodegenerative diseases.

Therefore, we developed a convolutional-neural-network-based model for predicting brain age based on raw SMRI images. We first processed the images so that all images are of the same scale, size, and orientation. Then we implemented a U-net feature extractor to automatically learn features from processed images. The features learned by the U-net model were used as input to a fully connected neural network to make brain-age predictions. We used the data of healthy individuals (input: raw SMRI images; output: brain age; ground truth: chronological age) to train the network and evaluated the model performance using the mean absolute error (MAE) between predicted brain ages and true ages in a test dataset containing more healthy individuals.

Subject Keywords: Brain Age; Convolutional Neural Network; Machine Learning; U-net; Image Processing; MRI images;

Acknowledgments

I would like to express my sincere appreciation and thank to my Ph.D advisor Yogatheesan Varatharajah. He is very patient, and his guidance is instructive. His explanation of the research and the existing work helped me significantly on starting my work. Every time I had questions and problems about the research, he replied to me so promptly that it encouraged me to continue my work more confidently.

I would also like to thank Professor Zbigniew T. Kalbarczyk for giving me a chance to work on this research. His suggestions on my presentation and thesis helped me to have a much clear and comprehensive understanding about research, presentations, and papers in general.

Data used in the preparation of this work were obtained from the MGH-USC Human Connectome Project (HCP) database (https://ida.loni.usc.edu/login.jsp). The HCP project (Principal Investigators : Bruce Rosen, M.D., Ph.D., Martinos Center at Massachusetts General Hospital; Arthur W. Toga, Ph.D., University of Southern California, Van J. Weeden, MD, Martinos Center at Massachusetts General Hospital) is supported by the National Institute of Dental and Craniofacial Research (NIDCR), the National Institute of Mental Health (NIMH), and the National Institute of Neurological Disorders and Stroke (NINDS). Collectively, the HCP is the result of efforts of co-investigators from the University of Southern California, Martinos Center for Biomedical Imaging at Massachusetts General Hospital (MGH), Washington University, and the University of Minnesota. [3]

Data used in the preparation of this thesis were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, M.D. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD). For up-to-date information, see www.adni-info.org. [1]

Data used in the preparation of this work were obtained from Information eXtraction from Images (IXI). More information can be obtained from https://brain-development.org/ixi-dataset/. [4]

Contents

1. Introduction

Brain age is a popular measure used in the study of brain aging that estimates the biological age of a brain based on the extent of cerebral atrophy. Accurate brain-age-prediction models have wide applicability in the clinical domain, e.g., predicting age-related neurodegenerative diseases. In general, brain age is estimated using structural magnetic resonance images (SMRI).

In order to achieve accurate predictions for arbitrary brain images, we trained machine learning models that take in 3D SMRI brain images as input and produce an estimation of brain age as output. A model that can produce accurate brain-age prediction can be applied to the clinical domain. There are many age-related neurodegenerative diseases like Alzheimer's disease. One important characteristic of these diseases is that they make a person's biological brain age older than the actual chronological age. For a healthy person, the biological brain age should be close to the chronological age. In contrast, for a person with Alzheimer's disease, the biological brain age would be much older than the chronological age. As indicated in [6], biological brain age for people with Alzheimer's disease would be about five years more than the chronological age. Therefore, the model that can accurately predict the brain age can be used to further predict possible diseases.

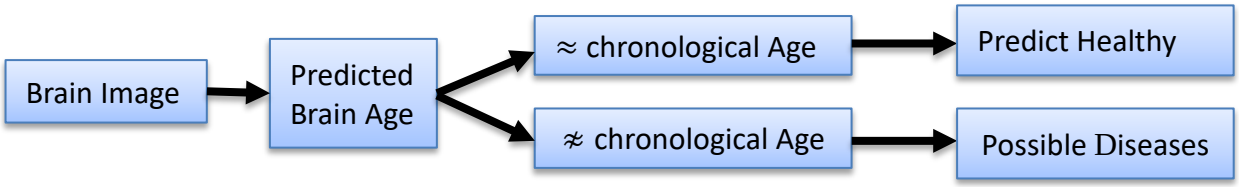

Figure 1 Usage of brain age prediction in clinical domain

2. Approach

The work flow for the brain-age prediction using SMRI can be divided into four main parts: ingestion, preprocessing, feature extractor, and regression. In the following, we describe each phase of the work flow, which has been implemented using TensorFlow [10]. TensorFlow is an application programming interface (API) in Python, which is designed for building up machine learning models.

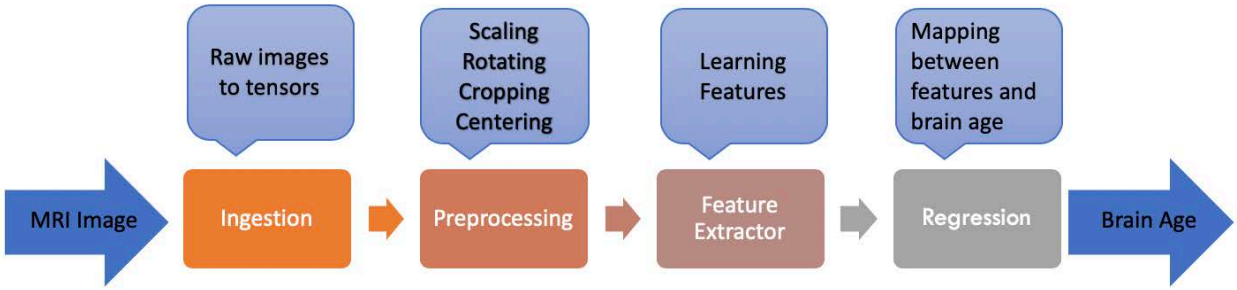

Figure 2 Work flow

2.1 Ingestion

In the ingestion phase, we read the raw MRI images and store the data into tensors. According to [10], a tensor is a generalization of vectors and matrices to potentially higher dimensions. The MRI images are stored in NIFTI format and come from three different data sources: the Alzheimer's Disease Neuroimaging Initiative (ADNI), the Information eXtraction from Images (IXI), and the Human Connectome Project (HCP). In our work, we choose to use healthy people's brain images as inputs to train the model, because healthy people's biological brain ages are close to their chronological ages. We can use the chronological ages of the healthy people to represent their brain ages as labels. The chronological age is just the difference between the date when the MRI image is taken and the date when the person is born.

2.2 Preprocessing

After reading the raw data and putting it into tensors, our work flow processes these tensors by applying image rotating, scaling, centering, and cropping to represent the tensors in the correct format.

We perform image rotating to make sure that the orientation of each axis of each image is the same. Since we have three different data sources, the layouts of the image data are different. Therefore, we have to rotate the images to align the same axis across all images. We use the original orientation of images read from the ADNI data source as the standard and do not perform rotations on them. For images read from the HCP data source, we rotate the images twice using the following code in Python:

```
im = rotate(im, 90, [0, 2])
im = rotate(im, 180, [1, 2])
```

Figure 3 Rotation for HCP images

For images read from the IXI data source, we rotate the images once using the following code in Python:

```
im = rotate(im, 90, [1, 0])
```

Figure 4 Rotation for IXI images

The following two images show the three-directional views for an example image from the ADNI data source without rotation, and the three-directional views for an example image from the IXI data source with rotation. The directions of the image layout are the same after all the rotations.

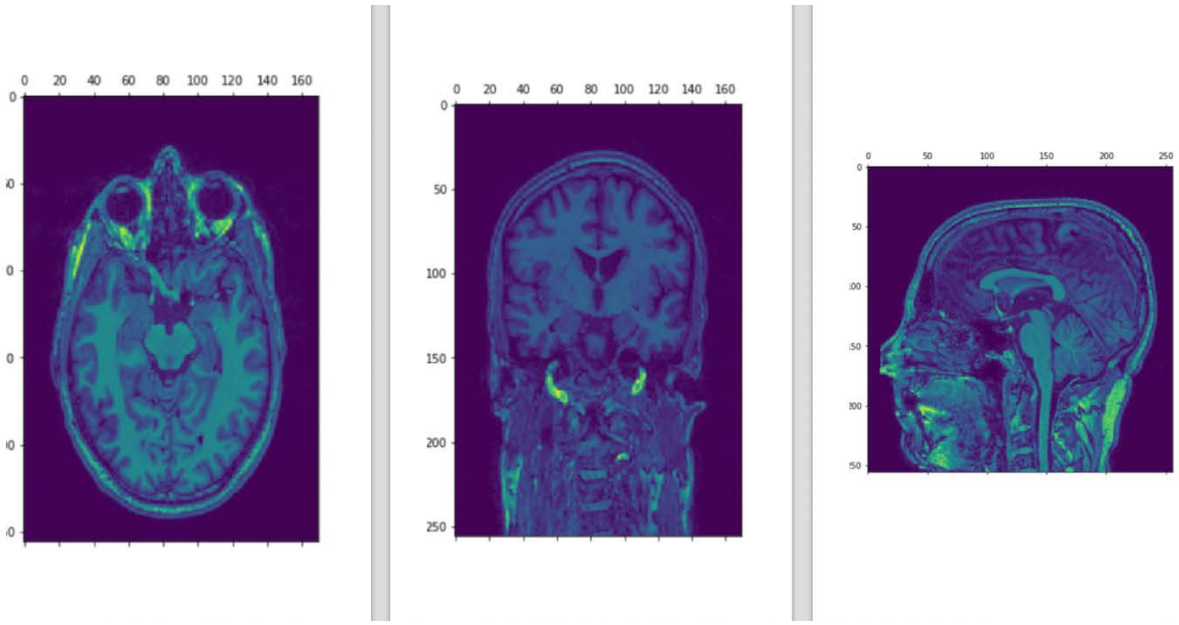

Figure 5 ADNI example image without rotation

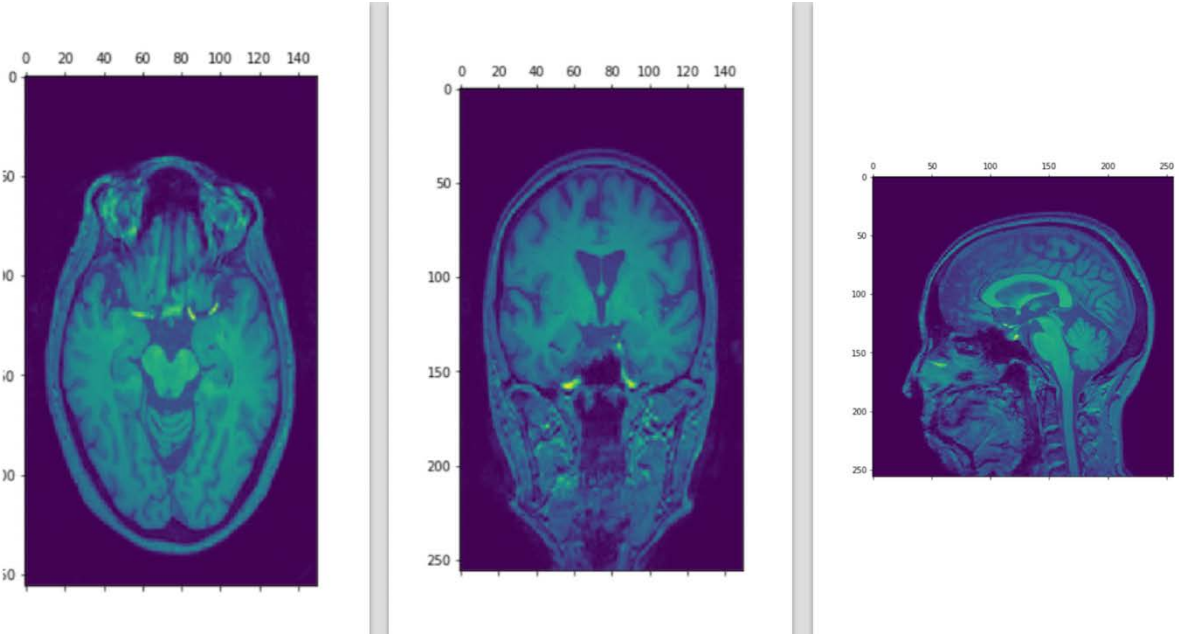

Figure 6 IXI example image with rotation

We perform image scaling to make sure that each work cell in each direction of each image has the same physical size when we feed the images into our machine learning model. Different from training normal data, if one image has different physical size from other images, the feature

extractor would not be able to get the accurate features because their physical locations are moved. Therefore, we need to read the physical size information from the NIFTI format files and rescale the images. Figure 7 and 8 show that the physical size of a pixel in the first direction of an example image from the HCP data source is different from the physical size of a pixel in the first direction of an example image from the ADNI data source.

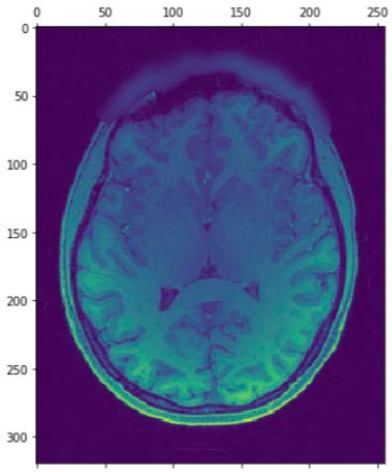

Figure 7 HCP example image in first direction

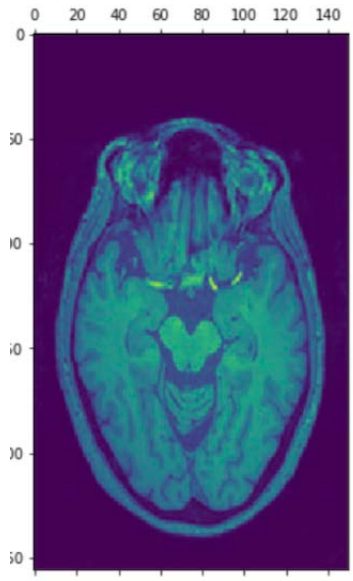

Figure 8 ADNI example image in first direction

We perform image centering to make sure that the physical locations of each part of brain of each image are similar. Different parts of brains serve different functionalities. For example, the cerebellum is responsible for movements and balance. In order to get accurate features, we want these specific parts to be located at similar places. However, some of the brain images are shifted from the center of the image. Therefore, we need to locate the center of the brain and align it with the center of the image.

We perform image cropping to make sure that each image has the same size when we fit it into the machine learning model.

2.3 Feature Extractor and Regression Model

The feature extractor and the regression are the machine learning models that we built. In our study, we compare the results of three existing models we found in the literature: Inception V1 discussed in [9], AlexNet discussed in [5], and U-net discussed in [7]. The Inception and the AlexNet can be directly used in our work because they would give a single number as output, which is the brain age. However, U-net is a convolutional neural network that is doing segmentation. It takes in an image as input and gives a scaled image with each pixel classified as output. We have to make modifications to U-net to make it work for our purpose. U-net is specifically designed for biomedical images, so after modifications we were hoping that it should give us an accurate prediction of the brain age.

U-net is consisted of 18 3x3 convolutional layers with batch normalization and rectified linear units (ReLU), four 2x2 max pooling, four 2x2 up convolutional layers, and one 1x1 convolutional layers. The first half is contracting the image. In each step in the first half, each

two 3x3 convolutional layers are followed by one 2x2 max pooling to down sample the image. Each time before the max pooling, the U-net crops part of the output and keeps it for the second half of the U-net. The first half consists of four such steps. In the second half of the U-net, in each step, it first does a 2x2 up convolution and then concatenates the cropped part from the first half to the layer output. This is followed by two 3x3 convolutional layers. The second half consists of four such steps and finally connects to the 1x1 convolutional layer to give the output.

As mentioned in [7], the U-net itself would not give us a prediction of brain age. It would produce a resized brain image with each pixel labeled with different classes. Therefore, we need to make modifications to the model to get the brain age as output. Since we do not rebuild the image, the second half, which is the up-sampling part, is not useful for our work. As a result, for our feature extractor, the first ten 3x3 convolutional layers and the first four max poolings are used. We choose to pick the output from one of the layers in these layers and feed the output to our regression model to produce the final result. The simple regression model we chose is just a fully connected layer with an output layer.

3. Literature Review

There are many papers discussing use of machine learning models to predict brain ages based on brain images. In [2], J. H. Cole, et al. built and trained a convolutional neural network that achieved a mean absolute error (MAE) of 4.16 years. The model consists of 3D convolutional layers, batch normalizations, rectified linear units (ReLU), max pooling, and fully connected layers. They are using the brain ages of 2001 healthy people aged 18 to 90 to train and test the model.

In [8], P. Sturmfels, et al. made two modifications to the model mentioned in [2]. Their model achieved a mean absolute error (MAE) of 1.6 years. The first modification made is regional segmentation. In different parts of human brain, different brain pattern means different things. Therefore, they separate the whole brain images into distinct regions. Each region has its own weights and biases when the model is trained and would not share its weights and biases with other regions. This will make sure each region keeps its region-specific information and further increase the accuracy of prediction. The second modification they made is about filter layout. Regular convolutional neural networks have small number of filters in convolutional layers in the beginning and increase the number of filters as more and more convolutional layers involved. However, this would cause the loss of region-specific information they get in the first modification. Therefore, they reversed the layout of the filter by having larger number of filters in convolutional layers in the beginning and decrease the number of filters as more and more convolutional layers involved. Because of this, their model focused more on information of different regions before they mixed together. They are using the brain ages of 1445 children with age from 8 to 21 to train and test their model.

Different from the previous two papers, our research focuses on other three machine learning models: Inception, AlexNet, and U-net. We built up our convolutional neural networks based on these three models and made modifications to them. In our study, we have a much larger data set of 3968 people's brain ages not limited to children.

4. Experiment Setup

In our study, the image data are classified into two different sets: training data and testing data. The training images are used to train our model to predict accurate brain ages, and the testing images are used to test whether our model is giving an accurate prediction based on the mean square error. We have 2292 images from the ADNI data source, 562 images from the IXI data source, and 1113 images from the HCP data source that are used for training. We have 3968 training images in total, and we have 199 testing images from the ADNI data source.

We first tested how the Inception and AlexNet models perform. Under the testing condition of 50,000 training steps and 0.00001 learning rate, the inception model gave a mean square error around 30 and the AlexNet model gave a mean square error around 300. The number of training steps stands for the number of times the models learn and adjust their weights and biases to achieve better result with less error. The learning rate influents the amount of adjustment the model would make during each training step. With a lower learning rate, the model would adjust its weights and biases with smaller values. For the inception model, a mean square error about 30 would produce an age error about 5.5 years. For the AlexNet model, a mean square error about 300 would produce an age error about 17.3 years.

For testing how the U-net performs, we start with 50,000 training steps and 0.00001 learning rate to test three models: two U-net layers with max pooling, four U-net layers with max pooling, and six U-net layers with max pooling. All of these models are connected with a fully connected layer and an output layer to give the final output. Each convolutional layer in the U-net is associated with a batch normalization. Under such circumstances, the two U-net layers model

and four U-net layers model gave us mean square errors around 2,000. The six U-net layers model gave us a much larger mean square error. However, all these three models were not showing a clear result because the mean square error was varying significantly from 2,000 to 4,000 before the training ended.

We referred to the AlexNet model which has a bias added after each convolutional layer instead of a batch normalization. We changed the batch normalization to add bias to see whether this will give us a better result. After being tested on these three models, although the mean square errors at the end are about the same, they gave clear and stable result without much variation.

With only 50,000 training steps, all the mean square errors of these models were still decreasing when the training ended. Therefore, we changed the training steps to 400,000 and kept the training rate the same as 0.00001 to see the final value that our model is converging to. For all the three models, all the mean square error finally converged to about 290. However, 290 is definitely not a value that can be called an accurate prediction because it gives an age error about 17 years. We thought the models gave such a big error because the learning rate was too big, so after a certain point, the model cannot be trained to make better prediction. Therefore, we modified code for changing the training rate. For the first 200,000 steps, we kept the training rate same at 0.00001. From 200,000 steps to 300,000 steps, the training rate was changed to 0.000001. For the last 100,000 steps, the training rate was changed to 0.0000001. However, for all the three models, the mean square errors still end up at around 290.

Since we cannot get a better result by changing how we train the model, we thought we might have to change our model. Because we do not have a large training and testing data set, the errors might come from overfitting. If overfit happens, we need to change our model to extract less features by decreasing the size of the layers we are using. In this case, we only use the first two layers of U-net since they would give us the less features. We decreased how many features each layer would extract by changing the size of layer from 64 to 8 or 16 or 32. The result shows that we are not getting a better result than before. The final mean square error is still around 290.

5. Description of Results

The final results are listed in Table 1, using 400,000 training steps and the modifying training rate.

Model	Mean Square Error	Estimated Age Error
Inception	27.5060	5.2440
AlexNet	305.9543	17.4915
2 U-net Layers	281.9081	16.7901
4 U-net Layers	285.9928	16.9113
6 U-net Layers	284.2462	16.8596
2 U-net Layers (Smaller layer size)	286.2589	16.9192

Table 1 Comparation Result

According to the results, we can see that changing the learning rate, the size of training steps, and the U-net structure would not help to decrease the mean square error of the prediction we get using U-net. Although the final mean square error we got using U-net structure is lower than the mean square error we got using the AlexNet model, the final age error around 16.8 years is too large to be considered as an accurate prediction. However, if we use the Inception model, we would get a mean square error around 27.5 which would give us an age error around 5.3 years. This is a much better than the U-net and AlexNet.

6. Conclusion

In this thesis, we introduced brain-age prediction research. We discussed papers about three different models: Inception, AlexNet and U-net. And we tested the performance of these models on predicting brain age using brain images. Especially for U-net, we introduced how to make modifications to it to predict brain age instead of doing segmentation. These models are not originally designed for brain-age prediction, so their performances are not as good as other models designed for brain-age prediction. However, this does not mean these models cannot be used for brain-age prediction. Especially, the Inception model gives an error close to the error given by the model in [2]. For future development, model modifications along the lines discussed in [8] can help to achieve more accurate prediction of the brain age.

References

[1] Alzheimer's Disease Neuroimaging Initiative (ADNI), www.adni-info.org. Accessed 16 Apr. 2019.

[2] J. H. Cole, et al. "Predicting Brain Age with Deep Learning from Raw Imaging Data Results in a Reliable and Heritable Biomarker." NeuroImage, vol. 163, pp. 115–124. December 2018.

[3] Human Connectome Project (HCP), https://ida.loni.usc.edu/login.jsp. Accessed 16 Apr. 2019.

[4] Information eXtraction from Images (IXI), https://brain-development.org/ixi-dataset/. Accessed 16 Apr. 2019.

[5] Alex Krizhevsky, et al. "ImageNet Classification with Deep Convolutional Neural Networks." Communications of the ACM, vol. 60, no. 6, pp. 84–90. June 2017.

[6] C. A. Raji, et al. "Age, Alzheimer Disease, and Brain Structure." Neurology, vol. 73, no. 22, pp. 1899–1905. December 2009.

[7] O. Ronneberger, et al. "U-Net: Convolutional Networks for Biomedical Image Segmentation." Vol. 9351, Springer Verlag. May 2015.

[8] P. Sturmfels, et al. “A Domain Guided CNN Architecture for Predicting Age from Structural Brain Images.” arXiv e-prints, arXiv:1808.04362. August 2018.

[9] C. Szegedy, et al. “Going Deeper with Convolutions.” Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 07-12-June-2015, pp. 1–9. September 2014.

[10] TensorFlow, https://www.tensorflow.org. Accessed 16 Apr. 2019.